SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Clayman, Gary L.

(ii) TITLE OF INVENTION: Methods and Compositions for the Diagnosis and Treatment of Cancer

(iii) NUMBER OF SEQUENCES: 14

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Arnold, White and Durkee
(B) STREET: P.O. Box 4433
(C) CITY: Houston
(D) STATE: TX
(E) COUNTRY: USA
(F) ZIP: 77210-4433

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: UNKNOWN
(B) FILING DATE: CONCURRENTLY HEREWITH
(C) CLASSIFICATION: UNKNOWN

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Highlander, Steven L.
(B) REGISTRATION NUMBER: 37,642
(C) REFERENCE/DOCKET NUMBER: INGN:022

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (512) 418-3000
(B) TELEFAX: (512) 474-7577

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2066 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CAAAACCTAC CAGGGCAGCT ACGGTTTCCG TCTGGGCTTC TTGCATTCTG GGACAGCCAA 60

GTCTGTGACT TGCACGTACT CCCCTGCCCT CAACAAGATG TTTTGCCAAC TGGCCAAGAC 120

CTGCCCTGTG CAGCTGTGGG TTGATTCCAC ACCCCCGCCC GGCACCCGCG TCCGCGCCAT 180

GGCCATCTAC AAGCAGTCAC AGCACATGAC GGAGGTTGTG AGGCGCTGCC CCCACCATGA 240

GCGCTGCTCA GATAGCGATG GTCTGGCCCC TCCTCAGCAT CTTATCCGAG TGGAAGGAAA 300

TTTGCGTGTG GAGTATTTGG ATGACAGAAA CACTTTTCGA CATAGTGTGG TGGTGCCCTA 360

TGAGCCGCCT GAGGTTGGCT CTGACTGTAC CACCATCCAC TACAACTACA TGTGTAACAG 420

TTCCTGCATG GGCGGCATGA ACCGGAGGCC CATCCTCACC ATCATCACAC TGGAAGACTC 480

CAGTGGTAAT CTACTGGGAC GGAACAGCTT TGAGGTGCGT GTTTGTGCCT GTCCTGGGAG 540

AGACCGGCGC ACAGAGGAAG AGAATCTCCG CAAGAAAGGG GAGCCTCACC ACGAGCTGCC 600

CCCAGGGAGC ACTAAGCGAG CACTGCCCAA CAACACCAGC TCCTCTCCCC AGCCAAAGAA 660

GAAACCACTG GATGGAGAAT ATTTCACCCT TCAGATCCGT GGGCGTGAGC GCTTCGAGAT 720

GTTCCGAGAG CTGAATGAGG CCTTGGAACT CAAGGATGCC CAGGCTGGGA AGGAGCCAGG 780

GGGGAGCAGG GCTCACTCCA GCCACCTGAA GTCCAAAAAG GGTCAGTCTA CCTCCCGCCA 840

TAAAAAACTC ATGTTCAAGA CAGAAGGGCC TGACTCAGAC TGACATTCTC CACTTCTTGT 900

TCCCCACTGA CAGCCTCCCA CCCCCATCTC TCCCTCCCCT GCGATTTTGG GTTTTGGGTC 960

TTTGAACCCT TGCTTGCAAT AGGTGTGCGT CAGAAGCACC CAGGACTTCC ATTTGCTTTG 1020

TCCCGGGGCT CCACTGAACA AGTTGGCCTG CACTGGTGTT TTGTTGTGGG GAGGAGGATG 1080

GGGAGTAGGA CATACCAGCT TAGATTTTAA GGTTTTTACT GTGAGGGATG TTTGGGAGAT 1140

GTAAGAAATG TTCTTGCAGT TAAGGGTTAG TTTACAATCA GCCACATTCT AGGTAGGGGC 1200

CCACTTCACC GTACTAACCA GGGAAGCTGT CCCTCACTGT TGAATTTTCT CTAACTTCAA 1260

GGCCCATATC TGTGAAATGC TGGCATTTGC ACCTACCTCA CAGAGTGCAT TGTGAGGGTT 1320

AATGAAATAA TGTACATCTG GCCTTGAAAC CACCTTTTAT TACATGGGGT CTAGAACTTG 1380

ACCCCCTTGA GGGTGCTTGT TCCCTCTCCC TGTTGGTCGG TGGGTTGGTA GTTTCTACAG 1440

TTGGGCAGCT GGTTAGGTAG AGGGAGTTGT CAAGTCTCTG CTGGCCCAGC CAAACCCTGT 1500

CTGACAACCT CTTGGTGAAC CTTAGATCCT AAAAGGAAAT GTCACCCCAT CCCACACCCT 1560

GGAGGATTTC ATCTCTTGTA TAGATGATCT GGATCCACCA AGACTTGTTT TAGCTCAGGG 1620

TCCAATTTCT TTTTTCTTTT TTTTTTTTTT TTTCTTTTTC TTTGAGACTG GGTCTCTTTG 1680

TTGCCCCAGG CTGGAGTGGA GTGGCGTGAT CTGGCTTACT GCAGCCTTTG CCTCCCCGGC 1740

TCGAGCAGTC CTGCCTCAGC CTCCGGAGTA GCTGGGACCA CAGGTTCATG CCACCATGGC 1800

CAGCCAACTT TTGCATGTTT TGTAGAGATG GGGTCTCACA GTGTTGCCCA GGCTGGTCTC 1860

AAACTCCTGG GCTCAGGCGA TCCACCTGTC TCAGCCTCCC AGAGTGCTGG GATTACAATT 1920

GTGAGCCACC ACGTCCAGCT GGAAGGGTCA ACATCTTTTA CATTCTGCAA GCACATCTGC 1980

ATTTTCACCC CACCCTTCCC CTCTTCTCCC TTTTTATATC CCATTTTTAT ATCGATCTCT 2040

TATTTTACAA TAAAACTTTG CTGCCA 2066

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Lys Thr Tyr Gln Gly Ser Tyr Gly Phe Arg Leu Gly Phe Leu His Ser
1 5 10 15

Gly Thr Ala Lys Ser Val Thr Cys Thr Tyr Ser Pro Ala Leu Asn Lys
20 25 30

Met Phe Cys Gln Leu Ala Lys Thr Cys Pro Val Gln Leu Trp Val Asp
35 40 45

Ser Thr Pro Pro Pro Gly Thr Arg Val Arg Ala Met Ala Ile Tyr Lys
50 55 60

Gln Ser Gln His Met Thr Glu Val Val Arg Arg Cys Pro His His Glu
65 70 75 80

Arg Cys Ser Asp Ser Asp Gly Leu Ala Pro Pro Gln His Leu Ile Arg
85 90 95

Val Glu Gly Asn Leu Arg Val Glu Tyr Leu Asp Asp Arg Asn Thr Phe
100 105 110

Arg His Ser Val Val Val Pro Tyr Glu Pro Pro Glu Val Gly Ser Asp
115 120 125

Cys Thr Thr Ile His Tyr Asn Tyr Met Cys Asn Ser Ser Cys Met Gly
130 135 140

Gly Met Asn Arg Arg Pro Ile Leu Thr Ile Ile Thr Leu Glu Asp Ser
145 150 155 160

Ser Gly Asn Leu Leu Gly Arg Asn Ser Phe Glu Val Arg Val Cys Ala

```
                    165                 170                 175

Cys Pro Gly Arg Asp Arg Arg Thr Glu Glu Glu Asn Leu Arg Lys Lys
             180                 185                 190

Gly Glu Pro His His Glu Leu Pro Pro Gly Ser Thr Lys Arg Ala Leu
         195                 200                 205

Pro Asn Asn Thr Ser Ser Ser Pro Gln Pro Lys Lys Lys Pro Leu Asp
     210                 215                 220

Gly Glu Tyr Phe Thr Leu Gln Ile Arg Gly Arg Glu Arg Phe Glu Met
225                 230                 235                 240

Phe Arg Glu Leu Asn Glu Ala Leu Glu Leu Lys Asp Ala Gln Ala Gly
                245                 250                 255

Lys Glu Pro Gly Gly Ser Arg Ala His Ser Ser His Leu Lys Ser Lys
             260                 265                 270

Lys Gly Gln Ser Thr Ser Arg His Lys Lys Leu Met Phe Lys Thr Glu
         275                 280                 285

Gly Pro Asp Ser Asp
     290
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2066 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CAAAACTTAC CAGGGCAACT ATGGCTTCCA CCTGGGCTTC CTGCAGTCTG GGACAGCCAA      60

GTCTGTTATG TGCACGTACT CTCCTCCCCT CAATAAGCTA TTCTGCCAGC TGGCGAAGAC     120

GTGCCCTGTG CAGTTGTGGG TCAGCGCCAC ACCTCCAGCT GGGAGCCGTG TCCGCGCCAT     180

GGCCATCCAC AAGAAGTCAC AGCACTTGAC GGGGGTCGTG AGACGCTGCC CCCACCATGA     240

GCGCTGCTCC GATGGTGATG GCCTGGCTCC TCCCCAGCAT CTTATCCGGG TGGAAGGAAA     300

TTTGTATCCC GAGTATCTGG AAGACAGGCA GACTTTTCGC CACAGCGTGG TGGTACCTTA     360

TGAGCCACCC GAGGCCGGCT CTGAGTATAC CACCATCCAC TACAAGTACA TTTGTAATAG     420

CTCCTGCATG GGGGGCATGA ACCGCCGACC TATCCTTACC ATCATCACAC TGGAAGACTC     480

CAGTGGGAAC CTTCTGGGAC GGGACAGCTT TGAGGTTCGT GTTTGTGCCT GCCCTGGGAG     540
```

AGACCGCCGT ACAGAAGAAG AAAATTTCCG CAAAAAGGAA GTCCTTTGCC CTGAACTGCC 600

CCCAGGGAGC GCAAAGAGAG CGCTGCCCAC CTGCACAAGC GCCTCTCCCC CGCAAAAGAA 660

AAAACCACTT GATGGAGAGT ATTTCACCCT CAAGATCCGC GGGCGTAAAC GCTTCGAGAT 720

GTTCCGGGAG CTGAATGAGG CCTTAGAGTT AAAGGATGCC CATGCTACAG AGGAGTCTGG 780

AGACAGCAGG GCTCACTCCA GCTACCTGAA GACCAAGAAG GGCCAGTCTA CTTCCCGCCA 840

TAAAAAAACA ATGGTCAAGA AAGTGGGGCC TGACTCAGAC TGACATTCTC CACTTCTTGT 900

TCCCCACTGA CAGCCTCCCA CCCCCATCTC TCCCTCCCCT GCCTTTTGGG TTTTGGGTCT 960

TTGAACCCTT GCTTGCAATA GGTGTGCGTC AGAAGCACCC AGGACTTCCA TTTGCTTTGT 1020

CCCGGGGCTC CACTGAACAA GTTGGCCTGC ACTGGTGTTT TGTTGTGGGG AGGAGGATGG 1080

GGAGTAGGAC ATACCAGCTT AGATTTTAAG GTTTTTACTG TGAGGGATGT TTGGGAGATG 1140

TAAGAAATGT TCTTGCAGTT AAGGGTTAGT TTACAATCAG CCACATTCTA GGTAGGGGCC 1200

CACTTCACCG TACTAACCAG GGAAGCTGTC CCTCACTGTT GAATTTTCTC TAACTTCAAG 1260

GCCCATATCT GTGAAATGCT GGCATTTGCA CCTACCTCAC AGAGTGCATT GTGAGGGTTA 1320

ATGAAATAAT GTACATCTGG CCTTGAAACC ACCTTTTATT ACATGGGGTC TAGATGACCC 1380

CCTTGAGGTG CTTGTTCCCT CTCCCTGTTG GTCGGTGGGT TGGTAGTTTC TACAGTTGGG 1440

CAGCTGGTTA GGTTGAGGTA GTTGTCAGGT CTCTGCTGGC CCAGCGAAAT TCTATCCAGC 1500

CAGTTGTTGG ACCCTGGCAC CTCAAATGAA ATCTCACCCT ACCCCACACC CTGTAAGATT 1560

CTATCTCTTG TATAGATGAT CTGGATCCAC CAAGACTTGT TTTAGCTCAG GGTCCAATTT 1620

CTTTTTTCTT TTTTTTTTTT TTTTTCTTTT TCTTTGAGAC TGGGTCTCTT TGTTGCCCCA 1680

GGCTGGAGTG GAGTGGCGTG ATCTGGCTTA CTGCAGCCTT TGCCTCCCCG GCTCGAGCAG 1740

TCCTGCCTCA GCCTCCGGAG TAGCTGGGAC CACAGGTTCA TGCCACCATG GCCAGCCAAC 1800

TTTTGCATGT TTTGTAGAGA TGGGGTCTCA CAGTGTTGCC CAGGCTGGTC TCAAACTCCT 1860

GGGCTCAGGC GATCCACCTG TCTCAGCCTC CCAGAGTGCT GGGATTACAA TTGTGAGCCA 1920

CCACGTCCAG CTGGAAGGGC CTACTTTCCT TCCATTCTGC AAAGCCCTGC TGCATTTATC 1980

CACCCCACCC TCCACCTGTC TCCCTCTTTT TTTCTTACCC CTTTTTATAT ATCAATTTCT 2040

TATTTTACAA TAAAATTTTG TTATCA 2066

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 293 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Lys Thr Tyr Gln Gly Asn Tyr Gly Phe His Leu Gly Phe Leu Gln Ser
1               5                  10                  15

Gly Thr Ala Lys Ser Val Met Cys Thr Tyr Ser Pro Pro Leu Asn Lys
            20                  25                  30

Leu Phe Cys Gln Leu Ala Lys Thr Cys Pro Val Gln Leu Trp Val Ser
        35                  40                  45

Ala Thr Pro Pro Ala Gly Ser Arg Val Arg Ala Met Ala Ile His Lys
    50                  55                  60

Lys Ser Gln His Met Thr Gly Val Val Arg Arg Cys Pro His His Glu
65                  70                  75                  80

Arg Cys Ser Asp Gly Asp Gly Leu Ala Pro Pro Gln His Leu Ile Arg
                85                  90                  95

Val Glu Gly Asn Leu Tyr Pro Glu Tyr Leu Glu Asp Arg Gln Thr Phe
            100                 105                 110

Arg His Ser Val Val Val Pro Tyr Glu Pro Pro Glu Ala Gly Ser Glu
        115                 120                 125

Tyr Thr Thr Ile His Tyr Lys Tyr Ile Cys Asn Ser Ser Cys Met Gly
    130                 135                 140

Gly Met Asn Arg Arg Pro Ile Leu Thr Ile Ile Thr Leu Glu Asp Ser
145                 150                 155                 160

Ser Gly Asn Leu Leu Gly Arg Asn Ser Phe Glu Val Arg Val Cys Ala
                165                 170                 175

Cys Pro Gly Arg Asp Arg Arg Thr Glu Glu Glu Asn Phe Arg Lys Lys
            180                 185                 190

Glu Val Leu Cys Pro Glu Leu Pro Pro Gly Ser Ala Lys Arg Ala Leu
        195                 200                 205

Pro Thr Cys Thr Ser Ala Ser Pro Pro Gln Lys Lys Lys Pro Leu Asp
    210                 215                 220

Gly Glu Tyr Phe Thr Leu Lys Ile Arg Gly Arg Leu Arg Phe Glu Met
225                 230                 235                 240
```

Phe Arg Glu Leu Asn Glu Ala Leu Glu Leu Lys Asp Ala His Ala Thr
245 250 255

Glu Glu Ser Gly Asp Ser Arg Ala His Ser Ser Tyr Leu Lys Ser Lys
260 265 270

Lys Gly Gln Ser Thr Ser Arg His Lys Lys Thr Met Val Lys Lys Val
275 280 285

Gly Pro Asp Ser Asp
290

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACTGCCCAAC AACACCA 17

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCCACGCCCA CACATTT 17

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCCTGTCCTG GGAGAGACCG 20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCCTTAAGCC ACGCCCACAC 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CACTGCCCAA CAACACCA 18

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GCCACGCCCA CACATTT 17

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGTGCATTGG AACGCGGATT 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGGACAGAA CGTTGTTTTC 20

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ACGGATTTGG TCGTATTGGG 20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TGATTTTGGA GGGATCTCGC 20